



ISRN TELCORDIA--2013-11+PR-0GARAU
Prepared for: Office of Naval Research

Fast Multiscale Algorithms for Information Representation and Fusion

Technical Progress Report No. 11

Devasis Bassu, Principal Investigator

Contract: N00014-10-C-0176

Applied Communication Sciences

150 Mount Airy Road

Basking Ridge, NJ 07920-2021

April 2013

Approved for public release; distribution is unlimited.

| Report Documentation Page | | | Form Approved OMB No. 0704-0188 | | |
|---|----------------|---|--|---|--|
| <p>Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> | | | | | |
| 1. REPORT DATE APR 2013 | 2. REPORT TYPE | 3. DATES COVERED 00-00-2013 to 00-00-2013 | | | |
| 4. TITLE AND SUBTITLE Fast Multiscale Algorithms for Information Representation and Fusion | | | 5a. CONTRACT NUMBER | | |
| | | | 5b. GRANT NUMBER | | |
| | | | 5c. PROGRAM ELEMENT NUMBER | | |
| 6. AUTHOR(S) | | | 5d. PROJECT NUMBER | | |
| | | | 5e. TASK NUMBER | | |
| | | | 5f. WORK UNIT NUMBER | | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Applied Communication Sciences,150 Mount Airy Road,Basking Ridge,NJ,07920 | | | 8. PERFORMING ORGANIZATION REPORT NUMBER | | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | | 10. SPONSOR/MONITOR'S ACRONYM(S) | | |
| | | | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) | | |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited | | | | | |
| 13. SUPPLEMENTARY NOTES | | | | | |
| 14. ABSTRACT | | | | | |
| 15. SUBJECT TERMS | | | | | |
| 16. SECURITY CLASSIFICATION OF: a. REPORT b. ABSTRACT c. THIS PAGE unclassified unclassified unclassified | | | 17. LIMITATION OF ABSTRACT Same as Report (SAR) | 18. NUMBER OF PAGES 12 | 19a. NAME OF RESPONSIBLE PERSON |

1 Abstract

In the eleventh quarter of the work effort, we focused on a) conducting experiments on real-world data sets using the developed algorithms, b) continued design/implementation of the Multiscale Heat-Kernel Coordinates (MHKC) algorithms with function estimation and c) packaging for releasing the software as open source. This report documents algorithm designs for the MHKC algorithms for visualization and function estimation.

The project is currently on track – in the upcoming final quarter, we will wrap up all pending work, write the final report and package all the deliverables. No problems are currently anticipated.

Table of Contents

| | | |
|-------|---|----|
| 1 | ABSTRACT | II |
| 2 | SUMMARY | 1 |
| 3 | INTRODUCTION | 2 |
| 4 | METHODS, ASSUMPTIONS AND PROCEDURES | 3 |
| 4.1 | Multiscale Heat Kernel Coordinates | 3 |
| 4.1.1 | Coordinate Charts | 3 |
| 4.1.2 | Function Estimation | 4 |
| 4.2 | Deliverables / Milestones | 5 |
| 5 | RESULTS AND DISCUSSION..... | 6 |
| 6 | CONCLUSIONS | 7 |
| 7 | REFERENCES | 8 |

2 Summary

In this quarter, we continued design and implementation of the new multiscale heat kernel coordinates (MHKC) algorithms for visualization and function extension. The design variants for MHKC algorithms are documented in this report.

The project is currently on track – in the upcoming final quarter, we will wrap up all pending work, write the final report and package all the deliverables. No problems are currently anticipated.

3 Introduction

The primary project effort over the last quarter focused on completing design/development of the multiscale heat-kernel coordinates algorithms [1]. An outline of the MHKC algorithms was presented in previous quarterly reports [10][11]. This report presents the MHKC algorithms for obtaining stable coordinate charts to represent the data in low dimensions. It also describes algorithms for performing function estimation for an arbitrary function defined on the data cloud.

4 Methods, Assumptions and Procedures

4.1 Multiscale Heat Kernel Coordinates

The Multiscale Heat Kernel Coordinates (MHKC) algorithms are based on theoretical results presented in [1]. The basic algorithms was described in [11]. We use the same notation in this report.

Given a set of n data points $\{x_1, x_2, \dots, x_n\}$ in R^d , we

1. Normalize the data by centering the cloud at zero and scaling to fit the data in a ball of unit variance,
2. Compute the transition probability matrix and its SVD

to obtain the eigenvalues $\{\lambda_i\}$ and associated eigenvectors $\{\vec{v}_i\}$. The MHKC embedding in r -dimensions for the point x_i is then given by

$$y_i = (e^{-\lambda_1 t} \cdot v_{i1}, e^{-\lambda_2 t} \cdot v_{i2}, \dots, e^{-\lambda_r t} \cdot v_{ir})$$

The heat kernel function is defined as

$$k(x, y) = \exp\left(-\frac{\|x - y\|_2^2}{\varepsilon}\right)$$

for any two points x and y . Here, ε is a constant (data dependent) representing the kernel window size.

4.1.1 Coordinate Charts

We now describe the algorithm to create canonical coordinate charts in 2 dimensions. This can be easily extended to r dimensions. The first step in to obtain a pair of points for each dimension. Next, we use these pairs as reference points to represent the entire data set. Specifically,

1. Center the 2-dimensional data cloud $\{y_i\}$ at the origin by subtracting the mean value
2. Find a pair of points $\{w_0, w_1\}$ in the dataset such that $(e^{-\lambda_1 t} \cdot v_1(w_0), e^{-\lambda_2 t} \cdot v_2(w_0))$ and $(e^{-\lambda_1 t} \cdot v_1(w_1), e^{-\lambda_2 t} \cdot v_2(w_1))$ are closest to the points $(1, -\frac{1}{4})$ and $(1, \frac{1}{4})$ respectively
3. Repeat step 2 to find another pair of points $\{z_0, z_1\}$ corresponding to the points $(-\frac{1}{4}, 1)$ and $(\frac{1}{4}, 1)$ respectively
4. The coordinate chart for the point x_i is then given by $(\log \frac{k(x_i, w_1)}{k(x_i, w_0)}, \log \frac{k(x_i, z_1)}{k(x_i, z_0)})$
5. Center the new coordinates at zero.

Note: The MSVD algorithm may be subsequently run on the coordinate chart to read off the local changes.

4.1.2 Function Estimation

Assume that we are given the function values $\{f(x_i)\}$ for an arbitrary unknown function f . The objective is to compute the value of $f(x)$ for a new point x that is not in the dataset. The function estimation algorithm is based on the Donoho-Johnstone technique for function estimation in the context of wavelets.

The function estimate at a new point x is given by

$$\hat{f}(x) = \frac{\sum_{i=1}^n k(x, x_i) \cdot f(x_i)}{\sum_{i=1}^n k(x, x_i)}$$

Note-1: The above equation is also a function of the window size ε . To obtain a stable function estimate, we require the estimate value to approximately remain the same across successive values of $\varepsilon = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$

Note-2: The above equation corresponds to a single time step in the diffusion process with the constraint that the random walk does not revisit the new point x . For arbitrary time t (keeping the constraint), we simply replace $k(x, x_i)$ by the i -th component of the vector given by

$$(k(x, x_1), k(x, x_2), \dots, k(x, x_n)) \cdot P^{t-1}$$

where P is the Markov transition probability matrix for the original dataset.

4.2 Deliverables / Milestones

| Date | Deliverables / Milestones | Status |
|----------|--|--------|
| Oct 2010 | Progress report for period 1, 1 st quarter | ✓ |
| Jan 2011 | Progress report for period 1, 2 nd quarter / complete randomized matrix decompositions task | ✓ |
| Apr 2011 | Progress report for period 1, 3 rd quarter / complete approximate nearest neighbors task | ✓ |
| Jul 2011 | Progress report for period 1, 4 th quarter / complete experiments – part 1 | ✓ |
| Oct 2011 | Progress report for period 2, 1 st quarter | ✓ |
| Jan 2012 | Progress report for period 2, 2 nd quarter / complete multiscale SVD task | ✓ |
| Apr 2012 | Progress report for period 2, 3 rd quarter | ✓ |
| Jul 2012 | Progress report for period 2, 4 th quarter / complete experiments – part 2 | ✓ |
| Oct 2012 | Progress report for period 3, 1 st quarter | ✓ |
| Jan 2013 | Progress report for period 3, 2 nd quarter / complete multiscale Heat Kernel task | ✓ |
| Apr 2013 | Progress report for period 3, 3 rd quarter | ✓ |
| Jul 2013 | Final project report + software + documentation on CDROM / complete experiments – part 3 | |

5 Results and Discussion

We described two additional MHKC algorithms – one to build stable canonical coordinate charts for representing the dataset, and the other to obtain function estimates for an unknown function defined on the dataset. We will experimentally evaluate both these techniques against real-world datasets.

6 Conclusions

The project is on track with design/implementation of the new multiscale heat kernel coordinates algorithms. We will finish up pending work, start writing up the final report and packaging all deliverables in the upcoming final quarter.

No problems are currently anticipated.

7 References

- [1] P.W. Jones, M. Maggioni, R. Schul, *Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels*, PNAS, vol. 105, no. 6, pp. 1803-1808.
- [2] E. Liberty, F. Woolfe, P.G. Martinsson, V. Rokhlin, M. Tygert, *Randomized Algorithms for the Low-Rank Approximation of Matrices*, PNAS, vol. 104, no. 51, pp. 20167-20172, 2007.
- [3] E. Liberty, F. Woolfe, P.G. Martinsson, V. Rokhlin, M. Tygert, *A Fast Randomized Algorithm for the Approximation of Matrices*, Applied and Computational Harmonic Analysis, vol. 25, pp.335-366, 2008.
- [4] V. Rokhlin, M. Tygert, A fast Randomized Algorithm for the Overdetermined Linear Least Squares Regression, PNAS, vol. 105, no. 36, pp. 13212-13217, 2008.
- [5] G. H. Golub, W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis, vol. 2 (2), pages 205–224, 1965.
- [6] G.H. Golub, C.F. Van Loan, *Matrix Computations (3rd edition.)*, Johns Hopkins University Press, Baltimore, 1996.
- [7] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion*, Technical Report No. 1, ISRN TELCORDIA--2010-01+PR-0GARAU, 2010.
- [8] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion*, Technical Report No. 2, ISRN TELCORDIA—2011-02+PR-0GARAU, 2011.
- [9] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion*, Technical Report No. 3, ISRN TELCORDIA—2011-03+PR-0GARAU, 2011.
- [10] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion*, Technical Report No. 9, ISRN TELCORDIA—2012-09+PR-0GARAU, 2012.
- [11] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion*, Technical Report No. 10, ISRN TELCORDIA—2013-10+PR-0GARAU, 2013.